Successful end to end generation of science products by ECS will involve the complex interaction of many of the ECS subsystems. Accurate testing of end to end system performance will require an integrated system testbed; this will not be available until late in the development period, well after integration of all of the system's software has begun.

As an intermediate step toward full system testing, the end to end generation of products can be viewed as a series of tightly coupled steps. Modeling and system analysis can be used to identify those steps which have the greatest associated technical risk. These high risk steps can then be studied separately as part of the risk reduction program.

Of greatest concern in the generation of products are the steps of data retrieval from the archive, data exchange between the Data Server and the Science Processing subsystems (DSS and DPS), and execution of the science software. The benchmarking plan defined below addresses several aspects of the latter two areas of concern, data exchange between DSS and DPS, and science software execution. Testing of archive system performance will be pursued under separate efforts.

SGI BDS, Filesystem I/O, and PGE Benchmarking Plan

1. Goals And Objectives

The goals of this effort are to validate design assumptions made for CDR and to gain early insight into PGE performance in the expected Release B hardware environment. These goals are detailed below.

1.1 BDS Performance

The CDR design calls for the use of a new software product, Bulk Data Service (BDS), offered by SGI as a lightweight TCP/IP protocol for transferring large amounts of data between systems at high speeds. The benchmarking effort will measure BDS throughput between systems equipped with filesystems and I/O configurations that are as similar as possible to those planned for the DAACs at Release B. The important measurements to be obtained are peak throughput (MB/second) for disk to disk transfers using a single HIPPI channel, and CPU utilization as a function of throughput. We are seeking to

validate our design assumption that BDS will provide a sustained throughput of 70MB/s over a single HIPPI channel, with minimal utilization (~25%) of a single processor.

1.2 Filesystem Performance

Filesystem performance -- disk I/O -- is an area of major concern for the Release B design because of the very high rates of disk access required by the Release B instruments, particularly MODIS. Our objective in performing these benchmarks is to implement filesystems as similar as possible to those that will be fielded for Release B, and to measure sustained throughput against these filesystems, using both I/O benchmarking software and science software. The filesystems under test will (if possible) be built using Irix 6.2 with XFS, striped over four SCSI-2 based RAID arrays, using the latest drives and controllers available from SGI. The CDR design for DPS assumes that we will be able to sustain 8 MB/second throughput from each SCSI-2/RAID channel; an objective of the effort is to validate this assumption.

1.3 PGE Performance

Performance of the actual science software is a major concern for the Release B design. If MODIS PGEs are available for test in the benchmark environment, data collected against a subset of the MODIS PGE suite would allow us to calibrate measured MODIS performance against expected performance; this would aid in future sizing estimates perpared for MODIS and other instruments. In addition (and time permitting), we will select one or more of the MODIS PGEs of particular interest and perform profiling and tuning of the code. The goal of this effort is not so much to provide tuned code back to the MODIS team, but rather to quantify the potential gains that may be realized through tuning, and to provide practical guidance (lessons learned) on tuning to all of the instrument teams.

The measurements that we will be taking will include CPU performance, I/O performance, and memory utilization for the PGEs. We will also try to examine resource utilization issues with several PGEs running at the same time -- i.e., contention for memory and disk resources.

2 Schedule

This effort is intended to support the development of the procurement package for the initial Release B buy -- that is, to help "right size" the Release B science processing hardware purchase. That procurement package is expected to be completed by the end of July. The schedule for this effort has therefore been constrained to complete (just) in time to support the procurement:

May 1	June 1	July 1	Aug 1
Complete equipment integration	**Complete software integration**	**Complete Testing**	**Complete Documentation**

3 Staffing Profile

The benchmarking effort is viewed as having two principal tasks, set-up/integration and measurement. The set-up and integration task will be performed with contributions from all of the interested parties -- the Multi-Release Support organization with Hughes ECS, the technical support organizations within Hughes ECS (HSTC and EDS), the MODIS SDST team, and SGI. Participation by the MODIS SDST and SGI groups is voluntary and uncompensated; however, without assistance from both groups in gathering equipment and integrating hardware and software, successful completion of the effort is doubtful.

The task of performing the benchmark measurements will be performed by the MRS group, with as much participation from MODIS SDST and SGI as they want to offer.

MRS Hardware LOE:

Joseph Blackette	50% April	80% May	80% June	20% July
Randy Miller	20% April	20% May	20% June	20% July

EDS/HTSC LOE:

40 hrs April	40 hrs May	40 hrs June	

MODIS SDST LOE:

40 hrs April	80 hrs May	80 hrs June

SGI LOE:

As Available/As Required

4 Equipment

4.1 Facility

The facility to be used for the test equipment is to be determined, and depends largely upon whose equipment is used. If MODIS SCF equipment is used extensively, it would

be expected that any additional equipment required would be installed there. It may be possible to use GSFC DAAC space within Building 32, or ECS space at the Hughes facility in Upper Marlboro.

4.2 Hardware

The table below shows the hardware desired for use in the tests.

Science Processor (SPRHW)	File System Management Server (FSMS)
CPU: 8 x 200 MHz R10000	CPU: 4 x 200 MHz R10000
RAM: 1GB/4-way interleaved	RAM: 512MB/4-way interleaved
OS: IRIX 6.2	OS: IRIX 6.2
IO4: 2	IO4: 2
HIO-1(1,1): FDDI	HIO-1(1,1): FDDI
HIO-2(1,2): HIPPI	HIO-2(1,2): HIPPI
HIO-3 (2,1): SCSI	HIO-3 (2,1): SCSI
HIO-4 (2,2): SCSI	HIO-4 (2,2): Unused
SCSI-0 (1,0,1): CD-ROM	SCSI-0 (1,0,1): CD-ROM
SCSI-1 (1,0,2): Two 4.3 GB Internal Disks	SCSI-1 (1,0,2): One 4.3 GB Internal Disk
SCSI-2 (2,0,1): RAID-1 SP1	SCSI-2 (2,0,1): RAID-5 SP1
SCSI-3 (2,0,2): RAID-1 SP2	SCSI-3 (2,0,2): RAID-5 SP2
SCSI-4 (2,1,1): RAID-2 SP1	SCSI-4 (2,1,1): RAID-6 SP1
SCSI-5 (2,1,2): RAID-2 SP2	SCSI-5 (2,1,2): RAID-6 SP2
SCSI-6 (2,1,3): RAID-3 SP1	
SCSI-7 (2,2,1): RAID-3 SP2	
SCSI-8 (2,2,2): RAID-4 SP1	
SCSI-9 (2,2,3): RAID-4 SP2	
RAID-1: 20 x 4.3GB	RAID-5: 20 x 4.3GB
RAID-2: 20 x 4.3GB	RAID-6: 20 x 4.3GB
RAID-3: 20 x 4.3GB	
RAID-4: 20 x 4.3GB	

As of this writing, none of this has been "acquired" for use in the tests.

Testing of BDS requires use of two machines connected via HiPPI. To accurately assess throughput, it is strongly desireable that these machines use Irix 6.2 and the corresponding version of XFS, which have been tuned to improve I/O performance. To assess CPU utilization by the HiPPI drivers, it is desireable that the systems be R10000-based. In order to test disk to disk transfers over the HiPPI, high speed filesystems (on the order of 70 MB/second) are required on both machines. This requires that each machine have at least four SCSI-2 paths to independently controlled disk arrays, so that four-way striping can be used. The DPS system has been configured with eight disk channels in order to support testing of PGE I/O in parallel with BDS, to observe the interaction of large and small request sizes on the filesystem.

Testing of MODIS PGEs is expected to be data-intensive; approximately 270 GB of space is desired on the DPS system, and approximately 135 GB on the FSMS system. It is highly desireable that the disk arrays reflect the newest technology from SGI (the Phoenix controller with Version 9 of the Flare code), which is to be used in Release B. Extrapolation of

performance from other disk arrays would be difficult. However, if SGI equipment is not available, use of other equipment would still permit some useful data on filesystem performance to be collected, and would allow the other objectives to be met.

In order to provide the most accurate estimate of PGE performance in the Release B environement, it is desireable to test with R10000 processors. However, shipment of these systems has been delayed, and it may be necessary to use R8000-based systems.

4.3 Software

4.3.1 Measurement Tools

We will use a combination of standard UNIX tools and a performance monitoring package from SGI. These are described below.

netstat. This Unix standard utility displays the HiPPI interface packet statistics in the of cumulative counts of input and output packets and errors. A unix "switch" inclusion changes netstat's behavior to the sampling of this information every second and it includes input and output byte counts.

osview. This SGI utility samples and displays various system-level statistics such as each CPU's utilization, memory utilization, and I/O buffer utilization.

top. This Unix tool provides a one-second interval of the top ten processes based on percent CPU utilization.

FORTRAN and C. We will be using SGI's MIPSpro compilers.

prof. We will be using SGI's standard profiling tool "prof" to help identify the most computationally complex section of PGE code.

iostat. A standard Unix utility that allows a user to measure throughput to a raw device.

dd. A standard Unix utility that allows transfer of variable sized data to and from file systems or raw devices at varying block sizes

xdd. A program developed to measure I/O peformance by reading and writing large amounts of data sequentially from a file or raw device. It is intended to find the upper limit of performance of an I/O subsystem under specific, well-controlled parameters.

Performance CoPilot. This will be included in the request to SGI. This provides insight into a number of performance statistics

4.3.2 Test Software

We will develop as little test software as possible. Test software -- mostly scripts -- will be used to drive the software under test (see below) and to collect and reduce data.

4.3.3 Software Under Test

The software under test includes a small set of MODIS PGEs and the BDS software. Working with the MODIS SDST team, we will identify a handful of MODIS PGEs that are available and interesting -- we will look for PGEs that are mature in their development, and use significant processing and/or I/O resources.

The BDS software will be requested from SGI.

5. Technical Plan

5.1 Equipment Acquisition, Installation, and Integration Plan

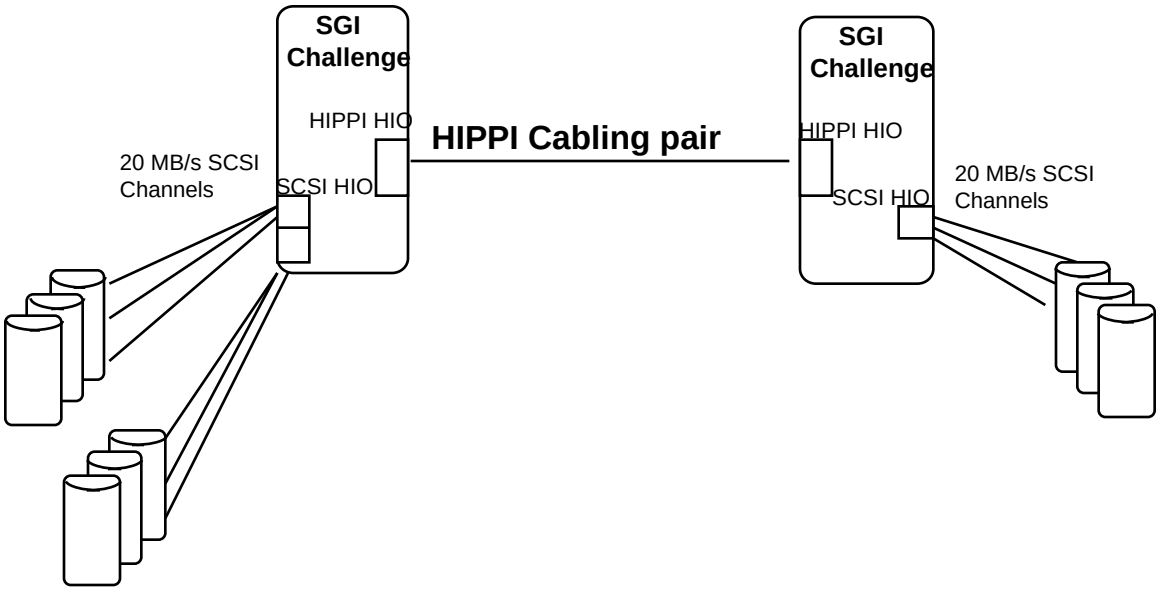
Plans for the acquisition, installation, and integration are largely to be determined. As a first step, we will meet with the MODIS SDST team to identify any hardware and facility resources that they may be able to make available. We will then contact SGI and request support from them.

5.2 Software Integration

The software integration task is principally one of becoming familiar with BDS, the test tools, and the MODIS PGEs. Some support from the MODIS team will be necessary to get their software running in a new environment. Whatever development of drivers, scripts, data collection, and data reduction software is required will be performed during this period.

5.3 Test Plans

Detailed test plans are to be developed.


SGI
Challenge
HIPPI HIO
SCSI HIO
20 MB/s SCSI
Channels
HIPPI Cabling pair
SGI
Challenge
HIPPI HIO
SCSI HIO
20 MB/s SCSI
Channels